TOWARDS MULTI-SCALE FEATURE DETECTION REPEATABLE OVER INTENSITY AND DEPTH IMAGES

Hatem A. Rashwan, Sylvie Chambon, Pierre Gurdjos, Géraldine Morin and Vincent Charvillat

University of Toulouse, CNRS-IRIT

ABSTRACT

Object recognition based on local features computed at multiple locations is robust to occlusions, strong viewpoint changes and object deformations. These features should be repeatable, precise and distinctive. We present an operator for repeatable feature detection on depth images (relative to 3D models) as well as 2D intensity images. The proposed detector is based on estimating the curviness saliency at multiple scales in each kind of image. We also propose quality measures that evaluate the repeatability of the features between depth and intensity images. The experiments show that the proposed detector outperforms both the most powerful, classical point detectors (e.g., SIFT) and edge detection techniques.

***Index Terms*—** Feature detectors, Curviness saliency, 2D-3D matching, Repeatability.

1. INTRODUCTION

A growing trend in recognition applications is to require 3D object models to overcome the limitations due to variations in viewpoint, texture or lighting that may modify the object appearance in the images. Since 3D capturing process is cheaper and faster, accurate 3D models become more available [1, 6, 17, 18]. In addition, 3D models can be integrated through separate dense depth images captured by range scanners (e.g., laser and IR) [4].

Recent approaches are based on 3D model databases [6]. These models can be represented by intermediate representations [11], like artificial images generated by rotating around the object through varying yaw, pitch and roll angles and focal length [4]. To cope with the shape variation independently of texture and lighting, an adapted representation is range (i.e., depth) images, which represent the object shape rather than its texture [6, 15]. The key features in both depth and intensity images are then detected to be matched. A key requirement on these features, as in 2D-2D matching, is to be computed with a high degree of *repeatability* (i.e., the probability that key features in the intensity image are found close to those extracted in the depth image must be high). In this paper, we require such assessments of repeatability, as shown in figure 1. Our objective is to introduce a detector robust to color, texture and illumination changes.

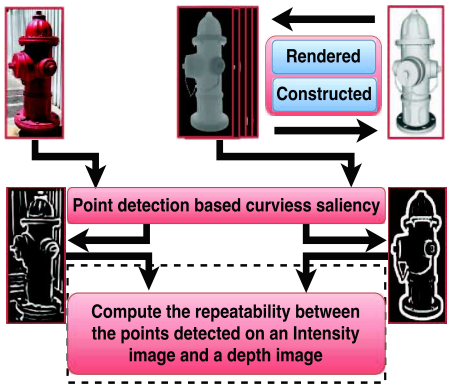


Fig. 1: The proposed detector based on curviness saliency to align intensity images to 3D models and the computation of the repeatability score between an intensity image and a corresponding depth image rendered from the same viewpoint.

To detect repeatable features in 2D, edges [5], corner detectors [10], eigenvalue analysis [16], multi-scale detectors (like SIFT [14] and SURF [2]) and curvature detection [9] are the four most important and the most used techniques. Recently, [8] presented curvature-based detector that use the structural cues to find the curvature in a multi-scale space. All these techniques are robust to lighting changes and translation; multi-scale approaches are also robust to scale and rotation. However, they depend on texture and/or color changes. Thus, the question is: what happens if we use these detectors for 2D-3D matching.

Recently, 2D-3D matching have been developed in the context of pose estimation. In [17], a sequence of silhouettes has been extracted from 3D models and input images. Then, shape similarity is measured between these silhouettes. In our paper, structural cues (e.g., curvilinear shapes) are extracted instead of only considering silhouettes since they are more robust to intensity, color, and pose variations. In fact, they not only represent outer contours (silhouette), but also inner (self-occluding) contours that also characterize the object. In addition, the histogram of gradients (HOG) detector [1, 13] or a fast version of HOG [6] have been also used to extract the features from rendering views and real images. All of these approaches give interesting results, however, they do not evaluate the repeatability between the set of points detected in an intensity image and those detected in an image rendered from

the 3D model. In fact, they all use a learning phase with manual matching but, in the context of our paper, we want to avoid a learning phase. Finally, in [18], the authors match the image with the 3D models by using SIFT in 2D and surface variation in 3D, but they assume that the object in the input image has no or poor internal texture.

As illustrated in figure 1, this work addresses the problem of aligning two images generated differently: an intensity image and a depth image that we assume to be taken from the same viewpoint. Curvature features do highlight geometric characteristics of an object. We propose a new detector based on curviness saliency that is a function of the eigenvalues of the Hessian matrix, an estimation of curvature. Furthermore, as relevant details exist only over a restricted range of scale, we further consider these features in a multi-scale analysis. This detection yields *more repeatable key points between intensity and depth images than the classical detectors*. Finally, an intensive evaluation and comparison have been conducted to highlight the quality of the proposed interest points.

2. PROPOSED MULTI-SCALE CURVINESS SALIENCY DETECTOR

Given an intensity image $\mathcal{I}$, we define the surface S such that $S(x,y) = (x, y, \mathcal{I}(x,y))$ under the assumption that $\mathcal{I}$ is twice differentiable. We denote by $\nabla\mathcal{I}$ the gradient vector of $\mathcal{I}$ and by $\mathsf{H} = \begin{pmatrix} \mathcal{I}_{xx} & \mathcal{I}_{xy} \\ \mathcal{I}_{xy} & \mathcal{I}_{yy} \end{pmatrix}$ the Hessian matrix of $\mathcal{I}$, i.e., the order-2 matrix of second-order partial derivatives $\mathcal{I}_{xx}$, $\mathcal{I}_{xy}$ and $\mathcal{I}_{yy}$. One key result is that the principal curvatures $\kappa_1(\mathbf{p})$, $\kappa_2(\mathbf{p})$ of S at point $\mathbf{p}$ are the eigenvalues of the matrix:

$$\widetilde{\mathsf{H}} \triangleq \alpha\,\mathsf{H}, \text{ where } \alpha = 1/\sqrt{1+\|\nabla\mathcal{I}\|^2}. \tag{1}$$

We aim at detecting "curvilinear features" in a representation common to depth and intensity images. By curvilinear features, we refer to points lying on elongated structures at which one principal curvature strongly dominates the other one. The Laplacian-energy is often used for image representation. So why not to use it here? Usually, the discrete form $\nabla^2\mathcal{I}(\mathbf{p}) = \mathcal{I}_{xx}(\mathbf{p}) + \mathcal{I}_{yy}(\mathbf{p}) \approx (1/h^2)\sum_{\mathbf{h}} \mathcal{I}(\mathbf{p}+\mathbf{h}) - \mathcal{I}(\mathbf{p})$, where $\mathbf{p} = (x,y)^\top$ and $\mathbf{h}$ varies in $\{\pm h\} \times \{\pm h\}$, is applied to each image and the obtained values are then squared. The problem is that the Laplacian operator clearly behaves like a mean rate of local intensity change so important directional information is lost. Linking this to curvatures, using the rotation invariance of the Laplacian, it can be shown that $\nabla^2\mathcal{I} = \alpha \operatorname{trace}\widetilde{\mathsf{H}} = \alpha(\kappa_1+\kappa_2)$ which means that $\nabla^2\mathcal{I}$ computes twice an "intensity-weighted" mean curvature of S.

The image representation proposed in this work is the so-called *curviness saliency* representation which relies on a function computing the difference between principal curvatures. We will now justify such a choice. Remind that, on the tangent plane T_S to S at point $\mathbf{p}$, for all unit directions $\mathbf{t}$ in T_S, (*i*) the normal curvatures $\kappa_{\mathbf{t}}(\mathbf{p})$ at $\mathbf{p}$ associated with $\mathbf{t}$ are the curvatures of the curves obtained by slicing S with the planes containing $\mathbf{p}$ and parallel to $\mathbf{t}$; (*ii*) $\rho_{\mathbf{t}}(\mathbf{q}) \triangleq 1/|\kappa_{\mathbf{t}}(\mathbf{q})|$ are the radii of curvature of these curves. A nice geometry result [3] is that all the points $\mathbf{q} = \mathbf{p} + \sqrt{\rho_{\mathbf{t}}(\mathbf{q})}\mathbf{t}$ on T_S, located at distance $\sqrt{\rho_{\mathbf{t}}}$ from $\mathbf{p}$, lie on a conic known as the Dupin indicatrix at $\mathbf{p}$. When $\mathbf{p}$ is the origin, the conic equation writes

$$(x,y)\widetilde{\mathsf{H}}(x,y)^\top = \pm 1. \tag{2}$$

Let choose a sign for $\pm\widetilde{\mathsf{H}}$ such that its eigenvalues be λ_1, λ_2 ensuring $\lambda_1 \geq \lambda_2$ and $\lambda_1 > 0$ (which are in fact equal to the two principal curvatures up to a common sign). Semi major- and minor-axes are $r_2 = |\lambda_2|^{-1/2}$ and $r_1 = |\lambda_1|^{-1/2}$ (since $r_2 \geq r_1$) respectively. The conic specializes to an ellipse if $\lambda_1\lambda_2 > 0$, or an hyperbola if $\lambda_1\lambda_2 < 0$ i.e., if λ_2 is negative.

The Dupin indicatrix yields a local information on the surface as the conic shape describes the 'distribution' of all normal curvatures at $\mathbf{p}$ (or more exactly of the squared roots of all radii of curvature). Various measures can describe this conic shape and we select the linear eccentricity $E\mp \triangleq \sqrt{r_2^2 \mp r_1^2}$, also called half-focal separation, which is the distance between the center and one focus, with '$-$' for ellipses and '$+$' for hyperpolas. Indeed, it can be easily shown that: $\sqrt{\lambda_1 - \lambda_2} = \frac{E\mp}{r_1 r_2}$, which provides a unified way of treating ellipses and hyperbolas (due to lack of space, the proof is omitted). The function:

$$\overline{\mathrm{CS}} \triangleq \lambda_1 - \lambda_2, \tag{3}$$

is large when $\lambda_1 \gg \lambda_2$, which means distant foci and so a highly elongated ellipse or a "squashed" hyperbola. This occurs e.g., when the point is located on a ridge (either curved or straight). In turn, when $\lambda_1 \simeq \lambda_2$, the conic approaches a circle and the distance between foci becomes very small. Therefore after computing (3) for every image pixel, any point with high value can be considered as a potential keypoint. Noting that the two eigenvalues $\lambda_1 \triangleq \lambda_+$ and $\lambda_2 \triangleq \lambda_-$ of the scaled Hessian matrix $\widetilde{\mathsf{H}}$ can be directly computed as:

$$\lambda_\pm = \frac{\alpha}{2}\left(\mathcal{I}_{xx} + \mathcal{I}_{yy} \pm \sqrt{(\mathcal{I}_{xx} - \mathcal{I}_{yy})^2 + 4\mathcal{I}_{xy}^2}\right), \tag{4}$$

the curviness saliency is then defined as

$$\mathrm{CS} \triangleq \overline{\mathrm{CS}}^2 = \alpha^2\left((\mathcal{I}_{xx} - \mathcal{I}_{yy})^2 + 4\mathcal{I}_{xy}^2\right). \tag{5}$$

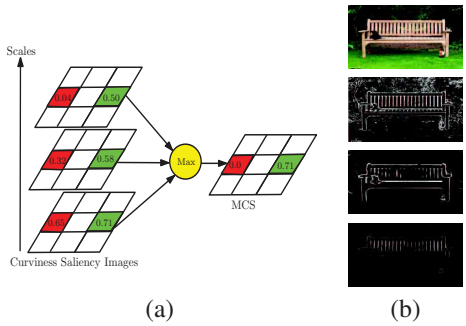
In figure 2, we show the different detections obtained using the minimum or the maximum eigenvalue, as proposed by Deng et al. [8]; the maximum eigenvalue provides a high response only for dark lines on a light background, while the minimum gives the high answer for the light lines on a dark background. Our proposition, the difference of the eigenvalues, improves robustness as it responds in both settings.

Computing the curviness saliency in a single-scale can only detect points that have high curvature in one scale and

Fig. 2: Curviness saliency of two shapes (columns 1, 5) with minimum (2, 6), maximum (3, 7) and the difference between maximum and minimum eigenvalues (4, 8).

high curvature points in other scales are missed. In consequence, in this paper, we compute the curviness saliency images in a multi-scale space. To build the scale pyramid, an edge-preserving smoothing approach, anisotropic diffusion filter [7] is used.

Contrary to depth images which represent textureless 3D shapes, intensity images are composed of shape and texture components. Consequently, the CS estimated from intensity images is affected by the textured regions. Our idea is to put forward the assumption that multi-scale analysis can discriminate between keypoints (those with high CS) due to shape and keypoints due to texture. As shown in figure 3.(b), at a coarse level, curves detected are reliable with poor localization and they miss small details. At a fine levels, details are preserved, but detection suffers greatly from clutters in textured regions. In addition, the CS values of small details and textures are high in the coarse level, whereas these values become lower in the finest levels. To combine the strengths of each scale, the CS value of each pixel over n scales is analyzed. If this value in all scales is higher than a threshold T, which is a function of the number of the smoothed images, m, (i.e., $T = e^{-m}$), the maximum curviness saliency (MCS) value of this pixel over all scales is then kept, see figure 3.(a). However, if the CS value is lower than T in one level, it is considered as texture (or small detail) point, thus it is removed from the final multiscale curviness saliency (MCS) image.



(a) (b)

Fig. 3: (a) A pyramid for a curviness saliency image computation: green values are over a threshold, so, the maximum is kept, whereas red values are under and replaced by 0. (b) Curviness saliency at different scales, top to bottom: the original image, the coarsest and finest levels, the final result.

3. EXPERIMENTS

To evaluate the repeatability of the detectors between intensity images and depth images, we apply a set of 9 detectors on a depth image, and its corresponding real images *taken from the same point of view*. The locations of features extracted on the depth images are considered as the ground-truth. Then, the extracted features on the real images are compared with the corresponding ground-truth. For all the tested 3D models, we generate depth images from approximately uniformly distributed viewing angles around a circle or a sphere[1].

Datasets – First, we use 2 real objects (i.e., stairs), with respectively 4 and 2 different textures, with their 3D CAD models. A set of depth images from different view is rendered and for each viewpoint, real images with different texture are captured, see figure 4. Moreover, we use 3D textureless objects (available online [2]), and we collect a set of 15 real images of each object on the web by choosing views as close as possible to the views used for the generation of the depth images. Moreover, to highlight the robustness of the approach to different acquisition conditions, many real images of a similar model are taken. Furthermore, we used the PASCAL3D+ dataset [19] that contains real images corresponding to 12 rigid objects categories; we compute average results for all non occluded objects in each category i.e., around 1000 objects per category. The real images are acquired under different acquisition conditions (e.g., lighting, complex background, low contrast). We rendered the depth image of the corresponding 3D CAD model using the viewpoint information from the dataset.

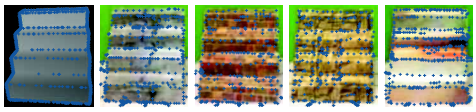


Fig. 4: Points detected with MCS on a depth image (column 1) and images with different textures rendered from the same viewpoint(column 2-5).

Comparison with existing detectors – The experiments include comparisons with these 9 following detectors[3]: *Edge-based detectors*: Sobel, Laplacian of Gaussian (Log), Canny [5] and Fuzzy logic technique [12], *corner detectors*: Harris detector, Minimum Eigenvalues detector [16], *multi-scale detectors*: SIFT, Scale Invariant Feature Transform [14], SURF, Speeded Up Robust Features [2], a multi-scale Principal Curvature Image (PCI) detector [8].

Setup – The images of the MCS results are calculated in scale space, similar to SIFT (see [7] for details about the construction of the pyramid). What is important is that we set the number of smoothed images per octave to 5, and in consequence, we have only 1 image result per octave.

[1] http://www.openu.ac.il/home/hassner/projects/poses/

[2] http://tf3dm.com/

[3] The proposed MCS detector is implemented in MATLAB. All tested point-like detectors and edge detection techniques are tested with the implementation given in MATLAB.

Methods	MCS		PCI		MinEig		Harris		SIFT		SURF		Sobel		Canny		LOG		Fuzzy	
Sequences	Ip	Hd	Ip	Hd	Ip	Hd	Ip	Hd	Ip	Hd	Ip	Hd	Ip	Hd	Ip	Hd	Ip	Hd	Ip	Hd
Stairs1	63	29	43	37	39	48	35	41	32	52	42	39	25	67	30	59	31	61	21	75
Stairs2	65	27	44	36	35	45	38	41	40	47	43	35	28	64	34	58	33	54	20	70
Car	50	29	46	40	08	57	04	77	03	85	03	71	10	48	18	46	11	47	05	49
Shoe	31	52	31	67	02	102	03	106	10	111	01	108	04	71	04	71	05	71	02	71
Plane	55	23	38	19	06	37	04	43	10	46	03	47	18	26	21	26	21	24	14	24
T-Rex	64	17	59	25	09	41	06	100	02	143	05	46	16	28	18	28	20	32	12	22
Elephant	32	41	32	55	03	80	03	91	05	114	03	74	06	57	08	58	06	57	04	57
Fhydrant	51	23	42	35	06	62	04	86	02	74	09	67	09	38	14	37	13	36	06	42
Jeep	62	31	58	42	05	70	05	67	05	74	06	89	09	47	15	47	11	46	06	47
Mug	54	56	50	65	02	129	03	133	04	134	03	145	08	72	12	76	07	75	08	75
Teddy	39	24	32	31	04	72	05	69	09	77	04	101	07	47	14	44	08	47	07	47
Pistol	67	16	61	26	09	34	09	96	09	44	04	73	13	30	23	65	14	29	07	26
plane	50	48	37	59	15	61	09	63	08	68	13	73	10	68	13	65	11	69	10	71
bicycle	61	75	57	79	25	90	08	101	16	93	24	100	13	83	15	84	18	82	14	87
boat	36	68	28	75	09	79	10	77	06	87	10	76	09	75	14	71	11	78	09	76
bus	24	110	17	117	05	128	06	123	02	131	04	127	04	121	06	118	04	122	04	123
car	41	85	24	98	08	102	08	100	03	113	06	108	16	89	18	88	14	94	13	97
chair	52	64	43	78	16	84	08	96	09	94	16	86	24	88	20	91	22	86	19	92
table	38	85	19	96	06	117	05	118	04	118	08	111	11	117	12	114	11	116	07	120
train	28	108	14	121	06	126	07	123	03	133	05	127	08	125	07	129	04	129	06	122
METime	0.018		0.041		0.022		0.057		0.121		0.088		0.023		0.024		0.062		0.176	

Table 1: Mean Intersection Percentage (IP) (*higher is better*) and Mean Hausdorff Distance (HD) (*lower is better*) of all depth image rendered from different viewpoints and all real images captured under different textures and lighting of, first, the two objects, second the 10 objects, and, third, the PASCAL3D+, with the proposed method (MCS) and 9 tested detectors. METime is the mean execution time in seconds of MATLAB codes of the MCS and 9 tested detectors executing on Intel Core(i7) 2.9 GHz.

Evaluation criteria – The behavior of the detectors is evaluated with 2 measures: *Hausdorff distance*, well known and used, and *Intersection percentage*, introduced in this paper and that evaluates the intersection between the set of points extracted from the 2 considered images. More precisely, for each point of the depth image, a corresponding point is sought at the same location in the real image. The intersection percentage (IP) counts the number of correspondences over the total number of points detected in the image.

Results – In figures 4 and 5, for the depth images, MCS gives points uniformly located on the silhouette of the objects and also some points inside the shape whereas the corner and the multi-scale detectors miss some specific parts of the shape, like the bottom of the car. Moreover, visually MCS performs best for detecting a set of points in the depth map close to those detected in the real images. All the detectors are affected by the background, i.e. they detect points that are not coherent with the depth map. For the edge operators, false contours are detected inside the car in the real images.

Fig. 5: Features extracted in the depth (row 1) and the real image (2) with: MCS (column 1), PCI proposed in [8] (2), SIFT (3), Harris (4) and Edge detector proposed in [12] (5).

As shown in Table 1, MCS is able to find the highest number of features in the intersection, with real images captured under different textures and lighting conditions. For real images, as background is arbitrary, a perfect score (100%) can not be expected. For the edge-detectors, the intersection percentage is less than 30% for all the edge operators. However, MCS still outperforms the other edge operators. Moreover, PCI, the closest approach to our, yields to good repeatability results that are comparable to MCS results. However, MCS still provides the best results among the 9 tested detectors. In addition, MCS provides the lowest *Hausdorff distance*. We also have to notice that PCI and Edge detectors perform better than the corner and the multi-scale detectors.

4. CONCLUSION

We introduce a feature detector, MCS, based on a multi-scale curviness saliency estimation that can extract points more repeatable than classical detectors when it is used between an intensity image and a depth image. We also proposed a quality measure, the intersection percentage, to evaluate the repeatability of the extracted features. The experiments show that MCS yields the best repeatability score. Future work aim at introducing this detector in a robust 2D-3D matching for robust object recognition. The next step will be to introduce a descriptor based on shape and common to these different images to increase the quality of the recognition. In addition, we aim at using defocus maps [20] to represent real images.

5. REFERENCES

[1] M. Aubry, D. Maturana, A. Efros, B. Russell and J. Sivic. Seeing 3D Chairs: Exemplar Part-Based 2D-3D Alignment Using A Large Dataset of CAD models. In CVPR. 2014.

[2] H. Bay, A. Ess, T. Tuytelaars and L. Van Gool. 2008. Speeded-Up Robust Features (SURF). CVIU. 110(3), 2008.

[3] W. Boehm and H. Prautzsch. Geometric Concepts for Geometric Design. CRC Press. 1993.

[4] R. J. Campbell and P. J. Flynn. A Survey Of Free-Form Object Representation and Recognition Techniques. CVIU. 81(2), 2001.

[5] J. Canny. A Computational Approach To Edge Detection. PAMI. 8(6), 1986.

[6] C. B. Choy, M. Stark, S. Corbett-Davies and S. Savarese. Enriching Object Detection with 2D-3D Registration and Continuous Viewpoint Estimation. In CVPR. 2015.

[7] S. Paris and F. Durand. 2009. A Fast Approximation of the Bilateral Filter Using a Signal Processing Approach. IJCV. 81(24–52), 2009.

[8] H. Deng, W. Zhang, E. Mortensen, T. Dietterich and L. Shapiro. Principal Curvature-Based Region Detector for Object Recognition. In CVPR. 2007.

[9] P. Fischer and T. Brox. Image Descriptors Based on Curvature Histograms. In German Conference on Pattern Recognition, GCPR. 2014.

[10] C. Harris and M. Stephens. A Combined Corner and Edge detector. In Alvey Vision Conference. 1988.

[11] A. Irschara, C. Zach, J.-M. Frahm and H. Bischof. From structure-from-motion point clouds to fast location recognition. In CVPR. 2009.

[12] E. Kiranpreet, K. Inderjeet, S. Gill and S. Firozpur. Fuzzy Logic Based Image Edge Detection Algorithm in MATLAB. International Journal of Computer Applications. 1(22). 2010.

[13] A. K. J. Lim and A. Torralba. Fpm: Fine Pose Parts-Based Model With 3D CAD Models. In ECCV. 2014.

[14] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV. 60(2) 2004.

[15] S. Song and J. Xiao. Sliding Shapes for 3D Object Detection in RGB-D Images. In ECCV. 2(4). 2014.

[16] J. Shi and C. Tomasi. Good Features to Track. In CVPR. 1994.

[17] A. Toshev, A. Makadia and K. Daniilidis. Shape-Based Object Recognition in Videos Using 3D Synthetic Object Models. In CVPR, 2009.

[18] Y. Y. Lee, M. K. Park, J. D. Yoo, and K. H. Lee. Multi-Scale Feature Matching Between 2D Image and 3D Model. SIGGRAPH Asia, 2013.

[19] Y. Xiang, R. Mottaghi and S. Savarese. Beyond PASCAL: A Benchmark for 3D Object Detection in the Wild. In WACV, 2014.

[20] S. Zhuo and T. Sim. Defocus Map Estimation from a Single Defocused Image. Pattern Recognition. 44(9). 2011.